

We Claim:

1. A method of distributing server load in an IP network, comprising:
building an association between a subscriber edge device and a
5 server controller using a packet switched network Quality of Service QoS
mechanism;
offering out a fair share of server bandwidth to the subscriber edge
device;
directing a resource request from a client using the subscriber edge
10 device to a server having an amount of the fair share of server bandwidth
required by the client, through a server controller; and
reserving the required fair share of server bandwidth, or a portion
thereof, for meeting the resource request of the client.
- 15 2. A method according to claim 1 further comprising:
releasing enough of the reserved share of bandwidth sufficient to allow
the network to admit a data flow of the requested resource on the remaining
share of reserved bandwidth.
- 20 3. A method according to claim 1 wherein said step of reserving
comprises:
using a portion of the available bandwidth for setting up a session for
admitting a data flow for the requested resource.
- 25 4. A method according to claim 1 wherein said step of reserving
comprises:
intercepting a reservation message issued by the server; and
keeping a tally on both the server controller and the subscriber edge
device of used bandwidth for a route through the IP network.
- 30 5. A method according to claim 1 wherein said step of reserving
comprises:

intercepting a reservation message issued by the server;

issuing a release message, ahead of the reservation message travelling from the server controller to the subscriber edge device, for releasing the requested amount of the offered fair share of server bandwidth;

5 and

forwarding the reservation message to the subscriber edge device behind the release message.

6. A method according to claim 1 further comprising:

10 determining the fair share of server bandwidth taking into account one of the following factors: capacity of the servers, capacity of the network, network connectivity of the servers, response time of the servers, distance to the servers, load on the servers, and fixed reservations for unsubscribed users.

15

7. A method according to claim 2 further comprising periodically recalculating the available bandwidth to each edge-based device.

8. A method according to claim 1 further comprising operating the server
20 controller to determine the fair share of the server bandwidth.

9. A method according to claim 1 further comprising:
providing a load balancing algorithm on the server controller; and
operating the load-balancing algorithm to determine the fair share of
25 server bandwidth.

10. A method according to claim 1 further comprising operating the edge-based device to select which fair share, or portion thereof, to reserve.

30 11. A method according to claim 1 further comprising:
providing a load balancing algorithm on the edge-based device; and

operating the load balancing algorithm to select which fair shares, or portion thereof, to reserve.

12. A method according to claim 1 further comprising:

5 providing a load balancing algorithm on each of the edge-based device and the server controller; and

operating the load balancing algorithms to manage the allocating, offering, and reserving of fair shares of server bandwidth for meeting the resource request of the client.

10

13. A method according to claim 1 wherein the QoS mechanism used is RSVP.

14. A method of distributing server load in an IP network, comprising:

15 building an association between a subscriber edge device and a server controller using a packet switched network Quality of Service QoS mechanism;

offering out a fair share of server bandwidth to the subscriber edge device;

20 requesting an amount of the offered fair share of server bandwidth from a server through the server controller;

intercepting a reservation message issued by the server for reserving the offered fair share of server bandwidth, or a portion thereof; and

25 keeping a tally on both the server controller and the subscriber edge device of used bandwidth for a route through the IP network.

15. A method according to claim 14 wherein the QoS mechanism used is RSVP.

30 16. A method of distributing server load in an IP network, comprising: building an association between a subscriber edge device and a

server controller using a packet switched network Quality of Service QoS mechanism;

offering out a fair share of server bandwidth to the subscriber edge device;

5 requesting an amount of the offered fair share of server bandwidth from a server through the server controller;

intercepting a reservation message issued by the server for reserving the requested amount of server bandwidth;

10 issuing a release message, ahead of the reservation message travelling from the server controller to the subscriber edge device, for releasing the requested amount of the offered fair share of server bandwidth; and

forwarding the reservation message to the subscriber edge device behind the release message.

15

17. A method according to claim 16 wherein the QoS mechanism used is RSVP.

18. A subscriber edge device for an IP-based network, comprising:

20 a data input for receiving data from the network;

a data output for sending data to the network;

25 a resource requester for identifying a request for a resource in data received at the data input and for sending a network request to a server controller associated with the source location of the resource specified in a client request;

a resource reserver for reserving an amount of bandwidth offered by the server controller through a QoS mechanism and for releasing any unneeded amount of bandwidth, wherein the device is adapted to make use of bandwidth optimally; and

30 a resource returner for obtaining the requested resource from the source location of the requested resource using the reserved amount of bandwidth, and for returning the requested resource to a client.

19. A device according to claim 18, wherein the resource reserver is configured to release an amount of unneeded bandwidth sufficient to admit a data flow of the requested resource on the remaining share of reserved bandwidth.
20. A device according to claim 18 further comprising, a load balancing algorithm for selecting which fair shares, or portion thereof, to reserve
21. A data center device for an IP-based network, comprising:
a data input for receiving data from the network;
a data output for sending data to the network;
a resource allocator for allocating fair shares of server bandwidth to the network; and
a server controller for offering the allocated fair shares of server bandwidth in response to a resource request received from the network, using a Quality of Service QoS mechanism.
22. A device according to claim 21 further comprising:
a load balancing algorithm for determining the fair share of server bandwidth.
23. A device according to claim 21, wherein the resource allocator comprises a load-balancing algorithm.
24. A device according to claim 21 further comprising:
a bandwidth calculator for periodically re-calculating the available bandwidth to the network.
25. A communications network comprising the subscriber edge device as claimed in claim 18.

26. A communications network comprising the data-center device as claimed in claim 21.

27. A communications network according to claim 26 further comprising a
5 subscriber edge device for an IP-based network, comprising:

a data input for receiving data from the network;

a data output for sending data to the network;

a resource requester for identifying a request for a resource in data
received at the data input and for sending a network request to a server
10 controller associated with the source location of the resource specified in a
client request;

a resource reserver for reserving an amount of bandwidth offered by
the server controller through a QoS mechanism and for releasing any
unnneeded amount of bandwidth, wherein the device is adapted to make use
15 of bandwidth optimally; and

a resource returner for obtaining the requested resource from the
source location of the requested resource using the reserved amount of
bandwidth, and for returning the requested resource to a client.

20 28. A computer readable storage medium storing instructions that, when
executed by a computer, cause the computer to perform a method of
distributing server load in an IP network, the method comprising:

building an association between a subscriber edge device and a
server controller using a packet switched network Quality of Service QoS
25 mechanism;

offering out a fair share of server bandwidth to the subscriber edge
device;

directing a resource request from a client using the subscriber edge
device to a server having an amount of the fair share of server bandwidth
30 required by the client, through the server controller; and

reserving the required fair share of server bandwidth, or a portion
thereof, for meeting the resource request of the client.

29. A computer readable storage medium storing instructions that, when executed by a computer, cause the computer to perform a method of distributing server load in an IP network, the method comprising:

building an association between a subscriber edge device and a
5 server controller using a packet switched network Quality of Service QoS mechanism;

offering out a fair share of server bandwidth to the subscriber edge device;

10 requesting an amount of the offered fair share of server bandwidth from a server through the server controller;

intercepting a reservation message issued by the server for reserving the offered fair share of server bandwidth, or a portion thereof; and

keeping a tally on both the server controller and the subscriber edge device of used bandwidth for a route through the IP network.

15

30. A computer readable storage medium storing instructions that, when executed by a computer, cause the computer to perform a method of distributing server load in an IP network, the method comprising:

building an association between a subscriber edge device and a
20 server controller using a packet switched network Quality of Service QoS mechanism;

offering out a fair share of server bandwidth to the subscriber edge device;

25 requesting an amount of the offered fair share of server bandwidth from a server through the server controller;

intercepting a reservation message issued by the server for reserving the requested amount of server bandwidth;

issuing a release message, ahead of the reservation message travelling from the server controller to the subscriber edge device, for

30

releasing the requested amount of the offered fair share of server bandwidth;
and

forwarding the reservation message to the subscriber edge device
behind the release message.